Functional Central Limit Theorems for Occupancies and Missing Mass Process in Infinite Urn Models

Journal of Theoretical Probability
https://doi.org/10.1007/s10959-020-01053-6

Functional Central Limit Theorems for Occupancies and Missing Mass Process in Infinite Urn Models

Mikhail Chebunin[1] · **Sergei Zuyev**[2]

Received: 26 June 2019 / Revised: 25 September 2020 / Accepted: 25 October 2020

Abstract

We study the infinite urn scheme when the balls are sequentially distributed over an infinite number of urns labeled 1,2,... so that the urn j at every draw gets a ball with probability p_j, where $\sum_j p_j = 1$. We prove functional central limit theorems for discrete time and the Poissonized version for the urn occupancies process, for the odd occupancy and for the missing mass processes extending the known non-functional central limit theorems.

Keywords Infinite urn scheme · Regular variation · Functional CLT · Occupancy process · Missing mass process

Mathematics Subject Classification (2020) 60F17 · 60G22 · 60G15 · 60G18

1 Introduction

In this paper, we study the following classical urn model first considered by Karlin [12]: $n \geq 1$ balls are distributed one by one over an infinite number of urns enumerated from 1 to infinity. The ball distributed at step $j = 1, 2 \ldots$, call it jth ball, gets into urn i with probability p_i, $\sum_{i=1}^{\infty} p_i = 1$, independently of the other balls. Such multinomial occupancy schemes arise in many different applications, in Biology [11], Computer science [13,14] and in many other areas, see, e.g., [10] and the references therein.

✉ Sergei Zuyev
sergei.zuyev@chalmers.se

Mikhail Chebunin
chebuninmikhail@gmail.com

1 Sobolev Institute of Mathematics SB RAS, Novosibirsk State University, Novosibirsk, Russia

2 Chalmers University of Technology, Gothenburg, Sweden

Published online: 23 November 2020

Let X_j be the urn the jth ball gets into and let $J_i(n)$ be the number of balls the ith urn contains after n balls are distributed:

$$J_i(n) = \sum_{j=1}^{n} \mathbb{I}_{X_j=i} \,.$$

Of a particular interest is the asymptotic behavior of the following quantities: the number of urns containing at least $k \geq 1$ balls and containing exactly k balls:

$$R^*_{n,k} = \sum_{i=1}^{\infty} \mathbb{I}_{J_i(n)\geq k}, \quad R_{n,k} = \sum_{i=1}^{\infty} \mathbb{I}_{J_i(n)=k} = R^*_{n,k} - R^*_{n,k+1}, \tag{1}$$

the number of urns with an odd number of balls and the *scaled missing mass* introduced in [12]:

$$U_n = \sum_{i=1}^{\infty} \mathbb{I}_{J_i(n)\equiv 1 \pmod 2}, \quad M_n = n \sum_{i=1}^{\infty} p_i \, \mathbb{I}_{J_i(n)=0}, \tag{2}$$

We also use notation $R_n \overset{\text{def}}{=} R^*_{n,1} = \sum_{k\geq 1} R_{n,k}$ for the number of non-empty urns. Renumbering the urns if necessary, we may assume that the sequence $(p_i)_{i\geq 1}$ is monotonely decaying. We further assume that it is regularly varying:

$$\alpha(x) = \max\{i \,:\, p_i \geq 1/x\} = x^{\theta} L(x) \text{ with} \theta \in [0, 1], \tag{3}$$

where $L(x)$ is a slowly varying function as $x \to \infty$.

Following Karlin's [12] original approach, we will consider a Poissonized version of the model when the balls are put into urns at the times of jumps of a homogeneous Poisson point processes $\Pi(s)$, $s \geq 0$ with intensity 1 on $\mathbb{R}_+$. According to the independent marking theorem for Poisson processes, $\{J_i(\Pi(s)) \overset{\text{def}}{=} \Pi_i(s),\ s \geq 0\}$ are independent homogeneous Poisson processes with intensities p_i. To ease the notation, we write simply

$$R(s) \overset{\text{def}}{=} R^*_{\Pi(s),1},\ U(s) \overset{\text{def}}{=} U_{\Pi(s)},$$

and we introduce the following Poissonized version of the scaled missing mass:

$$M(s) \overset{\text{def}}{=} s \sum_{i=1}^{\infty} p_i \, \mathbb{I}_{\Pi_i(s)=0} \,.$$

It differs from $M_{\Pi(s)}$ by the scaling factor s vs. $\Pi(s)$, but, when properly scaled, it is asymptotically equivalent to it.

Ordinary (not functional) central limit theorems for the above quantities were established under various conditions in [2,3,9,10,12–14]. In particular, under rather general conditions on the sequence (p_i) involving an unbounded growth of the variances, the following results are available: a strong law of large numbers and asymptotic normality of R_n, an asymptotic normality of the vector $(R_{n,1}, \dots, R_{n,\nu})$, local limit theorems, etc.

We acknowledge a novel method of a randomized decomposition for proving FCLTs developed in a recent paper [8], but we do not use it here. As a particular case of their Theorem 2.3, a FCLT holds for the processes R_n and U_n when $\theta \in (0, 1)$.

Our goal here is to establish a FCLT for the triplet of processes: the occupancy, odd occupancy and the scaled missing mass when $\theta \in (0, 1]$. In particular, we obtain previously unknown FCLT for U_n for $\theta = 1$ and for M_n when $\theta \in (0, 1]$. Up to a normalizing constant, the FCLT stated in Theorem 1 also holds for the original (non-scaled) missing mass $\sum_{i=1}^{\infty} p_i \, \mathbb{I}_{J_i(n)=0}$ on any interval $t \in [\varepsilon, 1]$, $\varepsilon > 0$, separated from 0. The paper extends the results of [6] and [7], where a functional central limit theorem (FCLT) was shown under condition (3) for the vector process $(R^*_{[nt],1}, R^*_{[nt],2}, \dots, R^*_{[nt],\nu})_{t\in[0,1]}$ in the case $\theta \in (0, 1]$.

Extending the FCLT to the case $\theta = 0$ would require additional to (3) conditions. As it was mentioned in [12] and in [2], $\theta = 0$ does not imply that the variances grow to infinity and various asymptotic behavior is possible for different statistics. We also argue that even an infinite growth of variances does not guarantee *per se* the required relative compactness.

When $\theta = 1$, we need a function

$$L^*(x) = \int_0^{\infty} L(xs) e^{-s} s^{-1} \mathrm{d}s.$$

It is known (see [12]) that $L^*(x)$ is slowly varying when $x \to \infty$.

Finally, for $t \in [0, 1]$ introduce the following notation:

$$\beta(n) = \begin{cases} \alpha(n), & \theta \in [0, 1); \\ nL^*(n), & \theta = 1, \end{cases} \qquad R_n(t) = \frac{R_{[nt]} - \mathbf{E}\, R_{[nt]}}{(\beta(n))^{1/2}}, \tag{4}$$

$$U_n(t) = \frac{U_{[nt]} - \mathbf{E}\, U_{[nt]}}{(\beta(n))^{1/2}}, \qquad M_n(t) = \frac{M_{[nt]} - \mathbf{E}\, M_{[nt]}}{(\alpha(n))^{1/2}}. \tag{5}$$

We are now ready to formulate the main result of the paper.

Theorem 1 *When $\theta \in (0, 1]$, the vector process*

$$(R_n(t), U_n(t), M_n(t)), \quad t \in [0, 1],$$

converges weakly in the uniform metric on $D([0, 1]^3)$ to a three-dimensional Gaussian process $(\rho(t), \upsilon(t), \mu(t))$ with zero mean and the covariance function $c(\tau, t)$ with the following components: when $\theta \in (0, 1)$, $\tau \le t$,

$$
\begin{aligned}
c_{\rho\rho}(\tau,t) &= \Gamma(1-\theta)((\tau+t)^\theta - t^\theta),\\
c_{\upsilon\upsilon}(\tau,t) &= \Gamma(1-\theta)2^{\theta-2}((t+\tau)^\theta - (t-\tau)^\theta),\\
c_{\mu\mu}(\tau,t) &= \theta\Gamma(2-\theta)\left(\frac{\tau}{t^{1-\theta}} - \frac{t\tau}{(t+\tau)^{2-\theta}}\right),\\
c_{\rho\upsilon}(\tau,t) &= \Gamma(1-\theta)((2t+\tau)^\theta - (2t-\tau)^\theta)/2,\\
c_{\rho\upsilon}(t,\tau) &= \Gamma(1-\theta)((2t+\tau)^\theta - t^\theta)/2,\\
c_{\rho\mu}(\tau,t) &= \theta\Gamma(1-\theta)\left(\frac{t}{(t+\tau)^{1-\theta}} - t^\theta\right),\\
c_{\rho\mu}(t,\tau) &= \theta\Gamma(1-\theta)\left(\frac{\tau}{(t+\tau)^{1-\theta}} - \frac{\tau}{t^{1-\theta}}\right),\\
c_{\mu\upsilon}(\tau,t) &= \theta\Gamma(1-\theta)\left(\frac{\tau}{2(2t+\tau)^{1-\theta}} - \frac{\tau}{2(2t-\tau)^{1-\theta}}\right),\\
c_{\mu\upsilon}(t,\tau) &= \theta\Gamma(1-\theta)\left(\frac{t}{2(2\tau+t)^{1-\theta}} - \frac{t^\theta}{2}\right).
\end{aligned}
$$

When $\theta = 1$, $\tau \le t$, $c(\tau,t)$ is given by

$$
\begin{aligned}
&c_{\rho\rho}(\tau,t) = \tau,\ c_{\upsilon\upsilon}(\tau,t) = 2\tau,\ c_{\mu\mu}(\tau,t) = \tau^2,\\
&c_{\rho\upsilon}(\tau,t) = \tau,\ c_{\rho\upsilon}(t,\tau) = (t+\tau)/2,\\
&c_{\rho\mu}(\tau,t) = c_{\rho\mu}(t,\tau) = c_{\upsilon\mu}(\tau,t) = c_{\upsilon\mu}(t,\tau) = 0.
\end{aligned}
$$

Thus, when $\theta = 1$, $\rho(t)$ and $\upsilon(t)$ are Wiener processes. For a general $\theta \in (0,1]$, the process $(\rho(t), \upsilon(t), \mu(t))$ is self-similar with the Hurst parameter $H = \theta/2$ which includes, in particular, a fractional Brownian motion, a bi-fractional Brownian motion with parameter $H = 1/2$, $K = \theta$ (see, e.g., [8]) with a new self-similar process $\mu(t)$.

2 Proof of Theorem 1

We start with formulating a couple of lemmas proved in [7]. We will generally use the letter C and its variants to denote a constant whose value is of no importance for us and note in parentheses the parameters it depends upon. This should not lead to a confusion when the same notation is used for, actually, different constants in different contexts, the same way $O(1)$ notation is used.

Lemma 1 *When $\theta > 0$, there exist $n_0 \ge 1$ and $C(\theta) < \infty$ such that*

$$\frac{\mathbf{E}\, R(n\delta)}{\beta(n)} \le C(\theta)\delta^{\theta/2}$$

holds for any $\delta \in [0,1]$ and $n \ge n_0$.

Lemma 2 *For any* $\varepsilon, \delta \in (0, 1)$ *there exists an* $N = N(\varepsilon, \delta)$ *such that for any* $n \ge N$,

$$\mathbf{P}(\forall t \in [0,1]\ \ \exists \tau : |\tau - t| \le \delta,\ \ \Pi(n\tau) = [nt]) \ge 1 - \varepsilon.$$

In preparation of the proof, let us introduce some further notation and establish a few inequalities we will be using.

In view of (5), let

$$U_n^*(t) = \frac{U(nt) - \mathbf{E}\,U(nt)}{(\beta(n))^{1/2}}, \qquad U_n^{**}(t) = \frac{U([nt]) - \mathbf{E}\,U([nt])}{(\beta(n))^{1/2}} \tag{6}$$

$$M_n^*(t) = \frac{M(nt) - \mathbf{E}\,M(nt)}{(\alpha(n))^{1/2}}, \qquad M_n^{**}(t) = \frac{M([nt]) - \mathbf{E}\,M([nt])}{(\alpha(n))^{1/2}}. \tag{7}$$

For any two positive $\tau_1 \le \tau_2$, define

$$\begin{aligned} U(\tau_2) - U(\tau_1) &= \sum_{i=1}^{\infty} \mathbb{I}\{\Pi_i(\tau_2) \text{ is odd}\} - \mathbb{I}\{\Pi_i(\tau_1) \text{ is odd}\} \\ &= \sum_{i=1}^{\infty} \mathbb{I}\{\Pi_i(\tau_2) \text{ is odd, } \Pi_i(\tau_1) \text{ is even}\} \\ &\quad - \mathbb{I}\{\Pi_i(\tau_2) \text{ is even, } \Pi_i(\tau_1) \text{ is odd}\} \\ &\stackrel{\text{def}}{=} \sum_{i=1}^{\infty} u_i(\tau_1, \tau_2) = \sum_{i=1}^{\infty} u_i = \sum_{i=1}^{\infty} u_i' - u_i'', \end{aligned}$$

and their expectations are denoted by

$$\overline{u}_i = \overline{u}_i' - \overline{u}_i'' = \overline{u}_i(\tau_1, \tau_2) \stackrel{\text{def}}{=} \mathbf{E}\,u_i' - \mathbf{E}\,u_i''.$$

Similarly for M,

$$\begin{aligned} M(\tau_2) - M(\tau_1) &= \sum_{i=1}^{\infty} (\tau_2 - \tau_1) p_i\ \mathbb{I}\{\Pi_i(\tau_2) = 0\} - \tau_1 p_i\ \mathbb{I}\{\Pi_i(\tau_1) = 0, \Pi_i(\tau_2) > 0\} \\ &\stackrel{\text{def}}{=} \sum_{i=1}^{\infty} m_i(\tau_1, \tau_2) = \sum_{i=1}^{\infty} m_i = \sum_{i=1}^{\infty} m_i' - m_i'', \\ \overline{m}_i &= \overline{m}_i' - \overline{m}_i'' = \overline{m}_i(\tau_1, \tau_2) \stackrel{\text{def}}{=} \mathbf{E}\,m_i' - \mathbf{E}\,m_i''. \end{aligned}$$

Clearly, for all natural k,

$$
\begin{aligned}
\mathbf{E}\,|u_i - \overline{u}_i|^k &= |1+\overline{u}_i|^k \overline{u}_i'' + |\overline{u}_i|^k (1-\overline{u}_i' - \overline{u}_i'') + |1-\overline{u}_i|^k \overline{u}_i' \\
&\le 2^k(\overline{u}_i' + \overline{u}_i'') + |\overline{u}_i|^k \le (2^k+1)(\overline{u}_i' + \overline{u}_i'') \\
&= (2^k+1)\left[\sum_{j=0}^{\infty} \mathbf{P}\{\Pi_i(\tau_1) = 2j,\ \Pi_i(\tau_2) - \Pi_i(\tau_1) \text{ is odd}\}\right. \\
&\quad \left. + \sum_{j=0}^{\infty} \mathbf{P}\{\Pi_i(\tau_1) = 2j+1,\ \Pi_i(\tau_2) - \Pi_i(\tau_1) \text{ is odd}\}\right] \\
&= (2^k+1)\mathbf{P}\{\Pi_i(\tau_2 - \tau_1) \text{ is odd}\} \\
&< (2^k+1)\mathbf{P}\{\Pi_i(\tau_2 - \tau_1) > 0\}. \qquad (8)
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
\mathbf{E}\,|m_i' - \overline{m}_i'|^k &\le 2^{k-1}(\mathbf{E}\,|m_i'|^k + |\overline{m}_i'|^k) = 2^{k-1}(\tau_2 - \tau_1)^k p_i^k (e^{-\tau_2 p_i} + e^{-k\tau_2 p_i}) \\
&< 2^k k!(1 - e^{-(\tau_2 - \tau_1)p_i}) = 2^k k!\,\mathbf{P}\{\Pi_i(\tau_2 - \tau_1) > 0\}, \\
\mathbf{E}\,|m_i'' - \overline{m}_i''|^k &\le 2^{k-1}(\mathbf{E}\,|m_i''|^k + |\overline{m}_i''|^k) < 2^k \tau_1^k p_i^k e^{-\tau_1 p_i}(1 - e^{-(\tau_2 - \tau_1)p_i}) \\
&< 2^k k!(1 - e^{-(\tau_2 - \tau_1)p_i}) = 2^k k!\,\mathbf{P}\{\Pi_i(\tau_2 - \tau_1) > 0\}.
\end{aligned}
$$

As a result,

$$
\mathbf{E}\,|m_i - \overline{m}_i|^k < 4^k k!\,\mathbf{P}\{\Pi_i(\tau_2 - \tau_1) > 0\}. \qquad (9)
$$

We are using the same notation $u_i,\ m_i$ and $\overline{u}_i,\ \overline{m}_i$ without explicitly specifying the corresponding values of $\tau_1 < \tau_2$; this should not create a confusion. The following lemma will be used in the proof of the relative compactness of the process $M_n^*(t)$.

Lemma 3 *Let $\theta \in (0, 1]$ and $\delta \in [0, 1]$. Then, there exist $n_0 \ge 1$ and $C(\theta) < \infty$ such that*

$$
\frac{\mathbf{var}(M(nt_2) - M(nt_1))}{\alpha(n)} \le C(\theta)\delta^{\theta/2}
$$

for all $t_2 - t_1 = \delta \ge 0$ and $n \ge n_0$.

Proof Put $\tau_2 = nt_2$ and $\tau_1 = nt_1$. Since the variance of an indicator does not exceed its expectation, we have that

$$\begin{aligned}
\mathbf{var}(M(\tau_2) - M(\tau_1)) &= \sum_{i=1}^{\infty} \mathbf{E}(m_i - \overline{m}_i)^2 = \sum_{i=1}^{\infty} \mathbf{E}(m_i')^2 - (\overline{m}_i' - \overline{m}_i'')^2 + \mathbf{E}(m_i'')^2 \\
&\le \sum_{i=1}^{\infty} (\tau_2 - \tau_1)^2 p_i^2 e^{-\tau_2 p_i} + \tau_1^2 p_i^2 e^{-\tau_1 p_i} (1 - e^{-(\tau_2-\tau_1)p_i} \pm (\tau_2 - \tau_1) p_i e^{-(\tau_2-\tau_1)p_i}) \\
&\le 2\frac{(\tau_2 - \tau_1)^2}{\tau_2^2} \mathbf{E}\, R_{\Pi(\tau_2),2} + \mathbf{E}\, R^*_{\Pi(\tau_2-\tau_1),2} + 6\frac{\tau_1^2(\tau_2 - \tau_1)}{\tau_2^3} \mathbf{E}\, R_{\Pi(\tau_2),3}.
\end{aligned}$$

By [12, Th. 2.1 and (23)],

$$\lim_{x\to\infty} \frac{\mathbf{E}\, R^*_{\Pi(x),2}}{\alpha(x)} = \Gamma(2-\theta) < 2,$$

and therefore, there exists an $x_1 > 1$ such that for all $x \ge x_1$,

$$\mathbf{E}\, R_{\Pi(x),2} + \mathbf{E}\, R_{\Pi(x),3} < \mathbf{E}\, R^*_{\Pi(x),2} < 2\alpha(x).$$

According to Karamata (see, e.g., [5, Th. 2,1, Eq. A6.2.10]), there exists an $x_2 > 0$ such that for all x and $\delta \in (0, 1]$ satisfying $x\delta \ge x_2$, one has

$$\frac{L(x\delta)}{L(x)} \le 2\delta^{-1/2}.$$

Let $n\delta > \max\{x_1, x_2\} = x_0$, then

$$\frac{\mathbf{E}\, R^*_{\Pi(n\delta),2}}{\alpha(n)} \le 2\frac{(n\delta)^\theta L(n\delta)}{n^\theta L(n)} \le 4\delta^{\theta/2}, \quad \frac{\max(\mathbf{E}\, R_{\Pi(nt_2),2}, \mathbf{E}\, R_{\Pi(nt_2),3})}{\alpha(n)} \le 4t_2^{\theta/2}.$$

Choose n_0 such that for all $n \ge n_0$ we have $n^\theta L(n) \ge n^{\theta/2}$. Then, provided $nt_2 \le x_0$,

$$\begin{aligned}
&\frac{\mathbf{E}\, R^*_{\Pi(n\delta),2}}{\alpha(n)} \le \frac{\mathbf{E}\, \Pi(n\delta)}{\alpha(n)} \le \frac{n\delta}{n^{\theta/2}} = (n\delta)^{1-\theta/2}\delta^{\theta/2} \le x_0\delta^{\theta/2}, \\
&\frac{\max(\mathbf{E}\, R_{\Pi(nt_2),2}, \mathbf{E}\, R_{\Pi(nt_2),3})}{\alpha(n)} \le x_0 t_2^{\theta/2}.
\end{aligned}$$

Now, take $c = \max\{4, x_0\}$. Since $t_2 - t_1 = \delta \ge 0$, for all $n \ge n_0$ we obtain

$$\frac{\mathbf{var}(M(nt_2) - M(nt_1))}{\alpha(n)} \le 2c\frac{\delta^2}{t_2^{2-\theta/2}} + \delta^{\theta/2} + 6c\frac{t_1^2\delta}{t_2^{3-\theta/2}} \le 9c \cdot \delta^{\theta/2}.$$

□

We are ready to prove Theorem 1. The proof is broken into four steps.
Step 1: Covariance The first rather technical step consists in establishing a formula for the covariances which is put in Appendix.
Step 2: Convergence of finite-dimensional distributions Along the lines of the proof of [9, Th. 12], one can show that for

$$m \geq 1, \quad 0 < t_1 < t_2 < \ldots < t_m \leq 1$$

the triangular array of m-dimensional vectors (i.e., independent in k for every n)

$$\left\{ \frac{\mathbb{I}(\Pi_k(nt_j) \text{ is odd}) - \mathbf{P}(\Pi_k(nt_j) \text{ is odd})}{\sqrt{\beta(n)}}, \; j \leq m, \; k \leq n \right\}_{n \geq 1}$$

satisfies the Lindeberg condition (see, e.g., [5, Th. 6.2]). Similarly, the convergence of the finite-dimensional distributions is shown for the process $M_n^*(t)$.
Step 3: Relative compactness

We shall follow the following plan:

(**a**) prove the continuity of the limiting process;
(**b**) prove that U_n^* and U_n^{**} (M_n^* and M_n^{**}) are sufficiently close;
(**c**) prove the relative compactness of U_n^{**} (M_n^{**}).

a(U) Take $\tau_1 = nt_1, \; \tau_2 = nt_2$ for $0 < t_1 < t_2 < 0$. Then,

$$\begin{aligned}
\mathbf{E}(U_n^*(t_2) - U_n^*(t_1))^2 &= \mathbf{E}\left(\sum_{i=1}^{\infty}(u_i - \overline{u}_i)\right)^2/\beta(n) = \sum_{i=1}^{\infty}\mathbf{E}(u_i - \overline{u}_i)^2/\beta(n) \\
&\leq 5\sum_{i=1}^{\infty}\mathbf{P}(\Pi_i(\tau_2 - \tau_1)>0)/\beta(n)=5\,\mathbf{E}\,R_{\Pi(\tau_2-\tau_1)}/\beta(n) \\
&\leq 5C(\theta)(t_2 - t_1)^{\theta/2}.
\end{aligned}$$

We have used above the independence of the summands, inequality (8) and Lemma 1.
Since the covariance function has a limit, [1, Th. 1.4] will imply that the limiting Gaussian process a.s. has a continuous modification on [0, 1].
Since the trajectories of the limiting Gaussian process belong a.s. to the class $C(0, 1)$, the weak convergence in the Skorohod topology implies the weak convergence in the uniform metric, see, e.g., [4]. Therefore, it is sufficient to prove the relative compactness of $\{U_n^*\}_{n \geq n_0}$ (with n_0 as in Lemma 1) in the Skorohod topology.
b(U) Since with probability one we have

$$|U(nt) - U([nt])| \leq \Pi(nt) - \Pi([nt]) \leq \Pi([nt] + 1) - \Pi([nt]),$$

then

$$\mathbf{E}\,|U(nt) - U([nt])| \leq 1.$$

Hence, for all $\eta > 0$,

$$\begin{aligned}
&\mathbf{P}(\sup_{0\le t\le 1} |U_n^*(t) - U_n^{**}(t)| > \eta) \\
&\quad \le \mathbf{P}(\sup_{0\le t\le 1} (|U(nt) - U([nt])| + \mathbf{E}\,|U(nt) - U([nt])|) > \eta\sqrt{\beta(n)}) \\
&\quad \le \mathbf{P}(\sup_{0\le t\le 1} (\Pi([nt]+1) - \Pi([nt]) + 1) > \eta\sqrt{\beta(n)}) \\
&\quad = \mathbf{P}(\sup_{0\le m\le n} (\Pi(m+1) - \Pi(m) + 1) > \eta\sqrt{\beta(n)}) \\
&\quad \le \sum_{m=0}^{n} \mathbf{P}(\Pi(m+1) - \Pi(m) + 1 > \eta\sqrt{\beta(n)}) \\
&\quad \le \sum_{m=0}^{n} \frac{\mathbf{E}\,e^{\Pi(m+1)-\Pi(m)+1}}{e^{\eta\sqrt{\beta(n)}}} = (n+1)\frac{\mathbf{E}\,e^{\Pi(1)}}{e^{\eta\sqrt{\beta(n)}-1}} = (n+1)e^{e-\eta\sqrt{\beta(n)}} \to 0
\end{aligned}$$

when $n \to \infty$. Therefore, it is sufficient to show the relative compactness of $\{U_n^{**}\}_{n\ge n_0}$ (with n_0 as in Lemma 1) in the Skorohod topology.
c(U) For any $t_1,\ t_2 \in [0,1]$ satisfying $\frac{1}{2n} \le t_2 - t_1$ we have that

$$\begin{aligned}
[nt_2] - [nt_1] &\le n(t_2 - t_1) + 1 \le n(t_2 - t_1) + 2n(t_2 - t_1) = 3n(t_2 - t_1) \\
&\le 3n(t_2 - t_1)\cdot(2n(t_2 - t_1))^3 = 24n^4(t_2 - t_1)^4. \qquad (10)
\end{aligned}$$

Put $k = [16/\theta] + 1$, $\tau_1 = [nt_1]$, $\tau_2 = [nt_2]$.

Recall the Rosenthal inequality [15]: if φ_i are independent random variables with $\mathbf{E}\,\varphi_i = 0$, then for all $k \ge 2$ there exists a constant $c(k)$ such that

$$\mathbf{E}\Big|\sum_i \varphi_i\Big|^k \le c(k)\max\Big\{\sum_i \mathbf{E}\,|\varphi_i|^k, \Big(\sum_i \mathbf{E}\,\varphi_i^2\Big)^{k/2}\Big\}. \qquad (11)$$

For all $n \ge n_0$ (with n_0 as in Lemma 1), we then have

$$\begin{aligned}
\mathbf{E}\,|U_n^{**}(t_2) - U_n^{**}(t_1)|^k &= \frac{\mathbf{E}\Big|\sum\limits_{i=1}^{\infty}(u_i - \overline{u}_i)\Big|^k}{(\beta(n))^{k/2}} \\
&\le \frac{c(k)}{(\beta(n))^{k/2}}\Big(\sum_{i=1}^{\infty}\mathbf{E}\,|u_i - \overline{u}_i|^k + \Big(\sum_{i=1}^{\infty}\mathbf{E}(u_i - \overline{u}_i)^2\Big)^{k/2}\Big) \\
&\le \frac{C(k)}{(\beta(n))^{k/2}}\Big(\sum_{i=1}^{\infty}\mathbf{P}(\Pi_i(\tau_2 - \tau_1) > 0) + \Big(\sum_{i=1}^{\infty}\mathbf{P}(\Pi_i(\tau_2 - \tau_1) > 0)\Big)^{k/2}\Big) \\
&= \frac{C(k)}{(\beta(n))^{k/2}}\Big(\mathbf{E}\,R(\tau_2 - \tau_1) + (\mathbf{E}\,R(\tau_2 - \tau_1))^{k/2}\Big)
\end{aligned}$$

$$\le \frac{C(k)}{(\beta(n))^{k/2}} \left(24n^4(t_2-t_1)^4 + (\mathbf{E}\, R(3n(t_2-t_1)))^{k/2}\right) \le \widetilde{C}(\theta)(t_2-t_1)^4,$$

where $c(k)$, $C(k)$ and $\widetilde{C}(\theta)$ depend only on their arguments.

Above, we have used (11) in the first inequality, (8) in the second and finally (10) and Lemma 1 alongside with the bound

$$\mathbf{E}\, R(\tau_2-\tau_1) \le \mathbf{E}(\Pi([nt_2]) - \Pi([nt_1])) = [nt_2] - [nt_1]. \tag{12}$$

If $0 \le t_2 - t_1 < \frac{1}{n}$, then $[nt_1] = [nt]$ or $[nt_2] = [nt]$ for all $t \in [t_1, t_2]$; therefore,

$$D \overset{\text{def}}{=} \mathbf{E}(|U_n^{**}(t) - U_n^{**}(t_1)|^{k/2} |U_n^{**}(t_2) - U_n^{**}(t)|^{k/2}) = 0 \le (t_2-t_1)^2.$$

If $t_2 - t_1 \ge 1/n$, then there are the following three cases:

1. if $t_2 - t \ge \frac{1}{2n}$, $t - t_1 \ge \frac{1}{2n}$, then the Cauchy–Schwarz inequality implies

$$D \le \widetilde{C}(\theta)(t_2-t)^2 \cdot (t-t_1)^2 \le \widetilde{C}(\theta)(t_2-t_1)^2.$$

2. If $t_2 - t \ge \frac{1}{2n}$, $t - t_1 < \frac{1}{2n}$, then since

$$|U([nt]) - U([nt_1])| \le_{\text{a.s.}} \Pi([nt]) - \Pi([nt_1]) \le_{st} \Pi(1),$$

the same inequality yields

$$D \le \left(\widetilde{C}(\theta)(t_2-t)^4 \cdot \mathbf{E}\left(\frac{\Pi(1)+1}{\sqrt{\beta(n)}}\right)^k\right)^{1/2} \le \widehat{C}(\theta)(t_2-t_1)^2.$$

3. If $t_2 - t < \frac{1}{2n}$, $t - t_1 \ge \frac{1}{2n}$, then since

$$|U([nt_2]) - U([nt])| \le_{\text{a.s.}} \Pi([nt_2]) - \Pi([nt]) \le_{st} \Pi(1),$$

we have that

$$D \le \left(\mathbf{E}\left(\frac{\Pi(1)+1}{\sqrt{\beta(n)}}\right)^k \cdot \widetilde{C}(\theta)(t-t_1)^4\right)^{1/2} \le \widehat{C}(\theta)(t_2-t_1)^2.$$

Now, the relative compactness follows from, for example, [4, Th. 13.5].

a(M) Because the covariance function has a limit, it is sufficient to appeal to Lemma 3 and [1, Th. 1.4] to establish existence of an almost sure continuous on [0, 1] modification of the limiting Gaussian process. Since the trajectories of this process are a.s. in $C(0,1)$, the weak convergence in the Skorohod topology implies the uniform convergence, see [4]. Thus, it is sufficient to prove a relative compactness of the family $\{M_n^*\}_{n\ge n_0}$ in the Skorohod topology (here, n_0 is the same as in Lemma 1).

b(M) Set $\tau_2 = nt$ and $\tau_1 = [nt]$. Since $\tau_2 - \tau_1 \leq 1$,

$$\begin{aligned}\mathbf{E}\,|M(\tau_2) - M(\tau_1)| \leq \sum_{i=1}^{\infty} &(\tau_2 - \tau_1) p_i e^{-p_i \tau_2} \\ &+ \tau_1 p_i e^{-p_i \tau_1}(1 - e^{-p_i(\tau_2 - \tau_1)}) \\ \leq \sum_{i=1}^{\infty} &p_i e^{-p_i \tau_2} + e^{-1} p_i(\tau_2 - \tau_1) < \sum_{i=1}^{\infty} 2p_i = 2.\end{aligned}$$

Let $m_i''' = m_i''(\tau_1, \tau_1 + 1)$ and $\overline{m}_i''' = \mathbf{E}\,m_i'''$. Then, we have almost surely

$$\begin{aligned}|M(\tau_2) - M(\tau_1)| &\leq \sum_{i=1}^{\infty}(m_i' + m_i'') \leq \sum_{i=1}^{\infty}(p_i + m_i''') \\ &= 1 + \sum_{i=1}^{\infty}(m_i''' + \overline{m}_i''' - \overline{m}_i''') < 2 + \left|\sum_{i=1}^{\infty}(m_i''' - \overline{m}_i''')\right|.\end{aligned}$$

We know that for any integer $k \geq 2$

$$\mathbf{E}\,|m_i''' - \mathbf{E}\,m_i'''|^k < 2^k k! \mathbf{P}(\Pi_i(\tau_1 + 1 - \tau_1) > 0) = 2^k k!(1 - e^{-p_i}) < 2^k k! p_i.$$

Using the independence of the terms and Rosenthal inequality, for any $k \geq 2$,

$$\begin{aligned}&\mathbf{E}\left|\sum_{i=1}^{\infty}(m_i''' - \overline{m}_i''')\right|^k \\ &\leq c(k)\left(\sum_{i=1}^{\infty}\mathbf{E}\,|m_i''' - \overline{m}_i'''|^k + \left(\sum_{i=1}^{\infty}\mathbf{E}(m_i''' - \overline{m}_i''')^2\right)^{k/2}\right) \\ &< c(k)(2^k k! + 4^k) = C(k).\end{aligned}$$

Hence, for $k \geq [2/\theta] + 1$ and all $\eta > 0$

$$\begin{aligned}&\mathbf{P}\left(\sup_{0\leq t\leq 1}|M_n^*(t) - M_n^{**}(t)| > \eta\right) \\ &\leq \mathbf{P}(\sup_{0\leq t\leq 1}(|M(nt) - M([nt])| + \mathbf{E}\,|M(nt) - M([nt])|) > \eta\sqrt{\alpha(n)}) \\ &\leq \mathbf{P}\left(\max_{0\leq[nt]\leq n}\left(\left|\sum_{i=1}^{\infty} m_i''' - \mathbf{E}\,m_i'''\right| + 4\right) > \eta\sqrt{\alpha(n)}\right) \\ &\leq \sum_{[nt]=m\in\{0,1,\ldots,n\}} \mathbf{P}\left(\left|\sum_{i=1}^{\infty} m_i''' - \mathbf{E}\,m_i'''\right| + 4 > \eta\sqrt{\alpha(n)}\right)\end{aligned}$$

$$\leq \sum_{m=0}^{n} \frac{C(k)}{(\eta\sqrt{\alpha(n)}-4)^k} = \frac{C(k)(n+1)}{(\eta\sqrt{\alpha(n)}-4)^k} \to 0 \text{ when } n \to \infty.$$

Therefore, it is sufficient to show the local compactness of $\{M_n^{**}\}_{n \geq n_0}$ in the Skorohod topology.

c(M) Let $t_1,\ t_2 \in [0,1]$ and $\frac{1}{2n} \leq t_2 - t_1$, then (10) holds. Set $k = [16/\theta] + 1$, $\tau_1 = [nt_1],\ \tau_2 = [nt_2]$.

Again, by independence and the Rosenthal inequality,

$$\begin{aligned}
\mathbf{E}\,|M_n^{**}(t_2) - M_n^{**}(t_1)|^k &= \frac{\mathbf{E}\left|\sum_{i=1}^{\infty}(m_i - \overline{m}_i)\right|^k}{(\alpha(n))^{k/2}} \\
&\leq \frac{c(k)}{(\alpha(n))^{k/2}}\left(\sum_{i=1}^{\infty}\mathbf{E}\,|m_i - \overline{m}_i|^k + \left(\sum_{i=1}^{\infty}\mathbf{E}(m_i - \overline{m}_i)^2\right)^{k/2}\right) \\
&\leq \frac{C(\beta)}{(\alpha(n))^{k/2}}\left(\sum_{i=1}^{\infty}\mathbf{P}(\Pi_i(\tau_2 - \tau_1) > 0) + (\mathbf{var}(M(\tau_2) - M(\tau_1)))^{k/2}\right) \\
&= \frac{C(k)}{(\alpha(n))^{k/2}}\left(\mathbf{E}\,R(\tau_2 - \tau_1) + (\mathbf{var}(M(\tau_2) - M(\tau_1)))^{k/2}\right) \\
&\leq \frac{C(k)}{(\alpha(n))^{k/2}}\left(24n^4(t_2 - t_1)^4 + (C(\theta)\alpha(n)(\tau_2 - \tau_1)/n)^{k/2}\right) \leq \widetilde{C}(\theta)(t_2 - t_1)^4,
\end{aligned}$$

where $c(k)$, $C(k)$ and $\widetilde{C}(\theta)$ depend only on their arguments.

Above, we have used inequalities (9), (10) and Lemmas 3, 1 alongside with the bound

$$\mathbf{E}\,R(\tau_2 - \tau_1) \leq \mathbf{E}(\Pi([nt_2] - [nt_1])) = [nt_2] - [nt_1].$$

When $0 \leq t_2 - t_1 < \frac{1}{n}$, then $[nt_1] = [nt]$ or $[nt_2] = [nt]$ for any $t \in [t_1, t_2]$. Thus,

$$B \stackrel{\text{def}}{=} \mathbf{E}(|M_n^{**}(t) - M_n^{**}(t_1)|^{k/2}|M_n^{**}(t_2) - M_n^{**}(t)|^{k/2}) = 0 \leq (t_2 - t_1)^2.$$

When $t_2 - t_1 \geq 1/n$, we have the following three cases:

1. if $t_2 - t \geq \frac{1}{2n}$, $t - t_1 \geq \frac{1}{2n}$, then the Cauchy–Schwarz inequality gives

$$B \leq \widetilde{C}(\theta)(t_2 - t)^2 \cdot (t - t_1)^2 \leq \widetilde{C}(\theta)(t_2 - t_1)^2;$$

2. if $t_2 - t \geq \frac{1}{2n}$, $t - t_1 < \frac{1}{2n}$, then since for any $l \geq 2$,

$$\begin{aligned}
&\mathbf{E}\,|M([nt]) - M([nt_1]) - \mathbf{E}(M([nt]) - M([nt_1])|^l \\
&\quad \leq \mathbf{E}\left(4 + \left|\sum_{i=1}^{\infty} m_i''([nt_1]+1, [nt_1]) - \mathbf{E}\,m_i''([nt_1]+1, [nt_1])\right|\right)^l < C(l),
\end{aligned}$$

the Cauchy–Schwarz inequality yields the bound

$$B \leq \left(\widetilde{C}(\theta)(t_2 - t)^4 \cdot \frac{C(k)}{\alpha(n)^{k/2}} \right)^{1/2} \leq \widehat{C}(\theta)(t_2 - t_1)^2;$$

3. finally, $t_2 - t < \frac{1}{2n}$, $t - t_1 \geq \frac{1}{2n}$, is similar to the previous case.

Thus, the required compactness follows from [4, Th. 13.5].

Finally, for the next step we need to show that $M(s)$, when time scaled, is close to its fully Poissonized version

$$\widetilde{M}(s) \stackrel{\text{def}}{=} M_{\Pi(s)} = \sum_{i=1}^{\infty} \Pi(s) p_i \ \mathbb{I}_{\Pi_i(s)=0} .$$

Namely, we aim to show that

$$\sup_{0 \leq t \leq 1} |M_n^*(t) - \widetilde{M}_n(t)| \to 0 \quad \text{in probability,} \tag{13}$$

where

$$\widetilde{M}_n(t) = \frac{\widetilde{M}(nt) - \mathbf{E}\,\widetilde{M}(nt)}{(\alpha(n))^{1/2}}.$$

Introduce $\Pi_i'(s) = \Pi(s) - \Pi_i(s)$ and $\widetilde{\Pi}(s) = (\Pi(s) - s)/\sqrt{s}$. Since $\widetilde{M}(s) = \sum_{i=1}^{\infty} \Pi_i'(s) p_i \ \mathbb{I}_{\Pi_i(s)=0}$,

$$\begin{aligned} |\,\mathbf{E}\,\widetilde{M}(s) - \mathbf{E}\,M(s)| &= |\,\mathbf{E} \sum_{i=1}^{\infty} (\Pi_i'(s) - s) p_i \ \mathbb{I}_{\Pi_i(s)=0}\,| \\ &= \left| \sum_{i=1}^{\infty} (s(1-p_i) - s) p_i e^{-sp_i} \right| = \frac{2\,\mathbf{E}\,R_{\Pi(s),2}}{s} \to 0 \end{aligned}$$

as $s \to \infty$ and it is bounded by 1. Thus, there exists a sufficiently small $\varepsilon = \varepsilon(\theta) > 0$ such that for $\delta_n = n^{\varepsilon - 1}$

$$\sup_{0 \leq t \leq \delta_n} |M_n^*(t) - \widetilde{M}_n(t)| < \frac{\Pi(n\delta_n) + n\delta_n + 1}{(\alpha(n))^{1/2}} \to 0 \text{ a.s.}$$

when $n \to \infty$.

By the strong law of large numbers for $M(s)$ and the well-known asymptotic behavior of $\mathbf{E}\,M(s)$ (see, e.g., [12][Eq. (23)]), we conclude that for any $\theta \in (0, 1]$, $M(s)/(s\alpha(s))^{1/2} \to 0$ a.s. when $s \to \infty$. Moreover, according to the central limit theorem $\widetilde{\Pi}(s)$ is asymptotically standard normal for large s.

Finally, we have almost surely

$$|M_n^*(t) - \widetilde{M}_n(t)| \le \frac{|\widetilde{\Pi}(nt)|M(nt)}{(nt\alpha(n))^{1/2}} + \frac{1}{(\alpha(n))^{1/2}}.$$

Using this inequality, and the fact that $\sup_{0\le t\le 1}(\cdot) \le \sup_{0\le t\le \delta_n}(\cdot) + \sup_{\delta_n\le t\le 1}(\cdot)$ and that $\sup\limits_{0\le t\le 1}(\cdot)$ is a continuous functional, we readily obtain 13.

Step 4: Approximation of the initial process Since $\Pi(t)$ is monotone, the strong law of large numbers implies that for any $\varepsilon, \delta \in (0, 1)$ there is an integer $N = N(\varepsilon, \delta)$ such that for all $n \ge N$ one has

$$\mathbf{P}(\forall t \in [0,1]\ \ \exists \tau : |\tau - t| \le \delta,\ \Pi(n\tau) = [nt]) \stackrel{\text{def}}{=} \mathbf{P}(A(n)) \ge 1 - \varepsilon,$$

see Lemma 2. Here and below, F stands for R, U or M. The relative compactness of the distributions $\{F_n^*\}_{n\ge n_0}$ implies that for any $\varepsilon \in (0, 1)$ and $\eta > 0$ there exist $\delta \in (0, 1)$ and an integer $N_1 = N_1(\varepsilon, \eta)$ such that for all $n \ge N_1$,

$$\mathbf{P}\left(\sup_{|t-\tau|\le\delta} \left|F_n^*(\tau) - F_n^*(t)\right| \ge \eta\right) \le \varepsilon.$$

Hence, since

$$\mathbf{P}(F_n(t) = F_n^*(\tau)|\Pi(n\tau) = [nt]) = 1,$$

for all $n \ge \max(N, N_1)$,

$$\begin{aligned}
&\mathbf{P}\left(\sup_{0\le t\le 1} \left|F_n(t) - F_n^*(t)\right| \ge \eta\right)\\
&\le \mathbf{P}\left(\sup_{0\le t\le 1} \left|F_n(t) - F_n^*(t)\right| \ge \eta,\, A(n)\right) + \varepsilon\\
&\le \mathbf{P}\left(\sup_{|t-\tau|\le\delta} \left|F_n^*(\tau) - F_n^*(t)\right| \ge \eta\right) + \varepsilon \le 2\varepsilon.
\end{aligned}$$

which proves Theorem 1.

Acknowledgements MC's research was supported by RSF Grant 17-11-01173-Ext. He also acknowledges hospitality of Chalmers University where a part of this work has been done. The authors are thankful to Sergey Foss for his interest in this research and valuable comments and to the anonymous reviewer for thorough reading and spotting some inaccuracies in the previous version of the manuscript.

Funding Open access funding provided by Chalmers University of Technology.

Appendix

An explicit expression for the covariance between $R(\tau)$ and $R(t)$ can be found in [7]. Take $\tau \le t$. The

$$
\begin{aligned}
c^*_{UU}(\tau, t) &= \mathbf{cov}(U(\tau), U(t)) \\
&= \sum_{k=1}^{\infty} \mathbf{P}(\Pi_k(\tau), \Pi_k(t) \text{ is odd}) - \mathbf{P}(\Pi_k(\tau) \text{ is odd})\mathbf{P}(\Pi_k(t) \text{ is odd}) \\
&= \frac{1}{4}\sum_{k=1}^{\infty} \Big((1 - e^{-2p_k\tau})(1 + e^{-2p_k(t-\tau)}) - (1 - e^{-2p_k\tau})(1 - e^{-2p_k t}) \Big) \\
&= \frac{1}{4}\sum_{k=1}^{\infty} e^{-2p_k(t-\tau)} - e^{-2p_k(t+\tau)} = \frac{1}{2}\mathbf{E}(U(t+\tau) - U(t-\tau)).
\end{aligned}
$$

Hence (since $\frac{\beta(nt)}{\beta(n)} \to t^\theta$ as $n \to \infty$)

$$
\begin{aligned}
c_{\upsilon\upsilon}(\tau, t) &= \lim_{n\to\infty} \frac{c^*_{UU}(n\tau, nt)}{\alpha(n)} = \Gamma(1-\theta)2^{\theta-2}((t+\tau)^\theta - (t-\tau)^\theta), \theta \in (0, 1), \\
c_{\upsilon\upsilon}(\tau, t) &= \lim_{n\to\infty} \frac{c^*_{UU}(n\tau, nt)}{nL^*(n)} = 2\tau, \ \theta = 1,
\end{aligned}
$$

cf. [12][Eq. (21)].

Next,

$$
\begin{aligned}
c^*_{MM}(\tau, t) &= \mathbf{cov}(M(\tau), M(t)) \\
&= \sum_{k=1}^{\infty} \mathbf{E}(tp_i\ \mathbb{I}(\Pi_i(t) = 0) - tp_i e^{-tp_i})(\tau p_i\ \mathbb{I}(\Pi_i(\tau) = 0) - \tau p_i e^{-\tau p_i}) \\
&= \sum_{k=1}^{\infty} t\tau p_i^2 e^{-tp_i}(1 - e^{-\tau p_i}) = \frac{2\tau}{t}\mathbf{E}\, R_{\Pi(t),2} - \frac{2t\tau}{(t+\tau)^2}\mathbf{E}\, R_{\Pi(t+\tau),2}.
\end{aligned}
$$

Since $\frac{\alpha(nt)}{\alpha(n)} \to t^\theta$ when $n \to \infty$,

$$c_{\mu\mu}(\tau, t) = \lim_{n\to\infty} \frac{c^*_{MM}(n\tau, nt)}{\alpha(n)}$$
$$= \theta\Gamma(2-\theta)\left(\frac{\tau}{t^{1-\theta}} - \frac{t\tau}{(t+\tau)^{2-\theta}}\right),$$

cf. [12][Eq. (23)].

Continuing,

$$\begin{aligned}
c^*_{RU}(\tau, t) &= \mathbf{cov}(R(\tau), U(t)) = \sum_{k=1}^{\infty} \mathbf{cov}(1 - \mathbb{I}(\Pi_k(\tau) = 0), \mathbb{I}(\Pi_k(t) \text{ is odd})) \\
&= \quad -\sum_{k=1}^{\infty} \mathbf{cov}(\mathbb{I}(\Pi_k(\tau) = 0), \mathbb{I}(\Pi_k(t) \text{ is odd})) \\
&= \quad -\sum_{k=1}^{\infty} \mathbf{P}(\Pi_k(\tau) = 0, \Pi_k(t) \text{ is odd}) - \mathbf{P}(\Pi_k(\tau) = 0)\mathbf{P}(\Pi_k(t) \text{ is odd}) \\
&= -\frac{1}{2}\sum_{k=1}^{\infty} \left(e^{-p_k\tau}(1 - e^{-2p_k(t-\tau)}) - e^{-p_k\tau}(1 - e^{-2p_k t})\right) \\
&= \frac{1}{2}\sum_{k=1}^{\infty} \left(e^{-p_k(2t-\tau)} - e^{-p_k(2t+\tau)} \pm 1\right) = \frac{1}{2}\mathbf{E}(R(2t+\tau) - R(2t-\tau)).
\end{aligned}$$

Similarly,

$$\begin{aligned}
c^*_{RU}(t, \tau) = \mathbf{cov}(R(t), U(\tau)) &= -\sum_{k=1}^{\infty} \mathbf{cov}(\mathbb{I}(\Pi_k(t) = 0), \mathbb{I}(\Pi_k(\tau) \text{ is odd})) \\
&= \frac{1}{2}\sum_{k=1}^{\infty} e^{-p_k t}(1 - e^{-2p_k\tau}) \\
&= \frac{1}{2}\sum_{k=1}^{\infty} \left(e^{-p_k t} - e^{-p_k(2\tau+t)} \pm 1\right) \\
&= \frac{1}{2}\mathbf{E}(R(2t+\tau) - R(t)).
\end{aligned}$$

Because $\frac{\beta(nt)}{\beta(n)} \to t^\theta$ when $n \to \infty$, for $\theta \in (0, 1)$ we have that

$$c_{\rho\upsilon}(\tau, t) = \lim_{n\to\infty} \frac{c^*_{RU}(n\tau, nt)}{\alpha(n)} = \Gamma(1-\theta)((2t+\tau)^\theta - (2t-\tau)^\theta)/2,$$
$$c_{\rho\upsilon}(t, \tau) = \lim_{n\to\infty} \frac{c^*_{RU}(nt, n\tau)}{\alpha(n)} = \Gamma(1-\theta)((2t+\tau)^\theta - t^\theta)/2.$$

For $\theta = 1$, this reduces to

$$c_{\rho\upsilon}(\tau, t) = \lim_{n\to\infty} \frac{c^*_{RU}(n\tau, nt)}{nL^*(n)} = \tau,$$
$$c_{\rho\upsilon}(t, \tau) = \lim_{n\to\infty} \frac{c^*_{RU}(nt, n\tau)}{nL^*(n)} = (t+\tau)/2,$$

cf. [12, Th. 1].

Next,

$$\begin{aligned} c^*_{MU}(\tau, t) &= \mathbf{cov}(M(\tau), U(t)) \\ &= \sum_{k=1}^{\infty} \tau p_k \, \mathbf{cov}(\mathbb{I}(\Pi_k(\tau) = 0), \mathbb{I}(\Pi_k(t) \text{ is odd})) \\ &= \frac{1}{2} \sum_{k=1}^{\infty} \tau p_k \left(e^{-p_k(2t+\tau)} - e^{-p_k(2t-\tau)} \right) \\ &= \frac{\tau}{2(2t+\tau)} \, \mathbf{E}\, M(2t+\tau) - \frac{\tau}{2(2t-\tau)} \, \mathbf{E}\, M(2t-\tau), \end{aligned}$$

and

$$\begin{aligned} c^*_{MU}(t, \tau) &= \mathbf{cov}(M(t), U(\tau)) \\ &= \frac{1}{2} \sum_{k=1}^{\infty} t p_k \left(e^{-p_k(2\tau+t)} - e^{-p_k t} \right) \\ &= \frac{t}{2(2\tau+t)} \, \mathbf{E}\, M(2\tau+t) - \frac{1}{2} \, \mathbf{E}\, M(t). \end{aligned}$$

Finally,

$$\begin{aligned} c^*_{RM}(\tau, t) &= \mathbf{cov}(R(\tau), M(t)) \\ &= \sum_{k=1}^{\infty} \mathbf{cov}(1 - \mathbb{I}(\Pi_k(\tau) = 0), t p_k \, \mathbb{I}(\Pi_k(t) = 0)) \\ &= -\sum_{k=1}^{\infty} t p_k \, \mathbf{cov}(\mathbb{I}\{\Pi_k(\tau) = 0\}, \mathbb{I}\{\Pi_k(t) = 0)\} \\ &= -\sum_{k=1}^{\infty} t p_k \left(e^{-p_k t} - e^{-p_k(\tau+t)} \right) = \frac{t}{\tau+t} \, \mathbf{E}\, M(\tau+t) - \mathbf{E}\, M(t), \end{aligned}$$

and

$$c^*_{RM}(t, \tau) = \mathbf{cov}(R(t), M(\tau)) = \frac{\tau}{\tau+t} \, \mathbf{E}\, M(\tau+t) - \frac{\tau}{t} \, \mathbf{E}\, M(t).$$

Because $\frac{\alpha(nt)}{\alpha(n)} \to t^\theta$ when $n \to \infty$, for $\theta \in (0,1)$ we obtain

$$\begin{aligned}
c_{\rho\mu}(\tau,t) &= \lim_{n\to\infty} \frac{c^*_{RM}(n\tau,nt)}{\alpha(n)} = \theta\Gamma(1-\theta)\left(\frac{t}{(t+\tau)^{1-\theta}} - t^\theta\right),\\
c_{\rho\mu}(t,\tau) &= \lim_{n\to\infty} \frac{c^*_{RM}(nt,n\tau)}{\alpha(n)} = \theta\Gamma(1-\theta)\left(\frac{\tau}{(t+\tau)^{1-\theta}} - \frac{\tau}{t^{1-\theta}}\right),\\
c_{\mu\upsilon}(\tau,t) &= \lim_{n\to\infty} \frac{c^*_{MU}(n\tau,nt)}{\alpha(n)} = \theta\Gamma(1-\theta)\left(\frac{\tau}{2(2t+\tau)^{1-\theta}} - \frac{\tau}{2(2t-\tau)^{1-\theta}}\right),\\
c_{\mu\upsilon}(t,\tau) &= \lim_{n\to\infty} \frac{c^*_{MU}(nt,n\tau)}{\alpha(n)} = \theta\Gamma(1-\theta)\left(\frac{t}{2(2\tau+t)^{1-\theta}} - \frac{t^\theta}{2}\right),
\end{aligned}$$

cf. [12][Eq. (23)].

Clearly, $L(n) \to 0$ as $n \to \infty$. According to [12][Lem. 4], in the case $\theta = 1$ the function $L^*(n) \to 0$ when $n \to \infty$ is slowly varying and

$$\lim_{n\to\infty} \frac{L(n)}{L^*(n)} \stackrel{\text{def}}{=} \lim_{n\to\infty} \delta_n = 0. \tag{14}$$

Therefore, in the case $\theta = 1$,

$$\begin{aligned}
c_{\rho\mu}(\tau,t) &= \lim_{n\to\infty} \frac{c^*_{RM}(n\tau,nt)}{\alpha(n)}\sqrt{\delta_n} = 0,\\
c_{\rho\mu}(t,\tau) &= \lim_{n\to\infty} \frac{c^*_{RM}(nt,n\tau)}{\alpha(n)}\sqrt{\delta_n} = 0,\\
c_{\mu\upsilon}(\tau,t) &= \lim_{n\to\infty} \frac{c^*_{MU}(n\tau,nt)}{\alpha(n)}\sqrt{\delta_n} = 0,\\
c_{\mu\upsilon}(t,\tau) &= \lim_{n\to\infty} \frac{c^*_{MU}(nt,n\tau)}{\alpha(n)}\sqrt{\delta_n} = 0.
\end{aligned}$$

References

1. Adler, R.J.: An Introduction to Continuity, Extrema, and Related Topics for General Gaussian Processes. Institute of Math. Stat, Hayward, California (1990)
2. Barbour, A.D., Gnedin, A.V.: Small counts in the infinite occupancy scheme. Electron. J. Probab. **14**(13), 365–384 (2009)
3. Ben-Hamou, A., Boucheron, S., Ohannessian, M.I.: Concentration inequalities in the infinite urn scheme for occupancy counts and the missing mass, with applications. Bernoulli **23**(1), 249–287 (2017)
4. Billingsley, P.: Convergence of Probability Measures, 2nd edition. Wiley, London (1999)
5. Borovkov, A.A.: Probability Theory. Universitext (2013)
6. Chebunin, M.G.: Functional central limit theorem in an infinite urn scheme for distributions with superheavy tails. Siberian Electron. Math. Rep. **14**, 1289–1298 (2017)
7. Chebunin, M.G., Kovalevskii, A.: Functional central limit theorems for certain statistics in an infinite urn scheme. Stats. Prob. Lett. **119**, 344–348 (2016)
8. Durieu, O., Wang, Y.: From infinite urn schemes to decompositions of self-similar Gaussian processes. Electron. J. Prob. **21**(43), 1–23 (2016)

9. Dutko, M.: Central limit theorems for infinite urn models. Ann. Probab. **17**, 1255–1263 (1989)
10. Gnedin, A., Hansen, B., Pitman, J.: Notes on the occupancy problem with infinitely many boxes: general asymptotics and power laws. Probab. Surv. **4**, 146–171 (2007)
11. Good, I.J., Toulmin, G.H.: The number of new species, and the increase in population coverage, when a sample is increased. Biometrika **43**(1/2), 45–63 (1956)
12. Karlin, S.: Central limit theorems for certain infinite urn schemes. J. Math. Mech. **17**(4), 373–401 (1967)
13. Muratov, A., Zuyev, S.: Bit flipping and time to recover. J. Appl. Prob. **53**(3), 1–17 (2016)
14. Orlitsky, A., Santhanam, N., Zhang, J.: Universal compression of memoryless sources over unknown alphabets. IEEE Trans. Inf. Theory **50**(7), 1469–1481 (2004)
15. Rosenthal, H.P.: On the subspaces of $l_p (p>2)$ spanned by sequences of independent random variables. Israel J. Math. **8**(3), 273–303 (1970)